

ORACLE

AI/ML for Customer

企業志向のAI実用化

西井 雄飛

AIエバンジェリスト
日本オラクル株式会社

Safe harbor statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

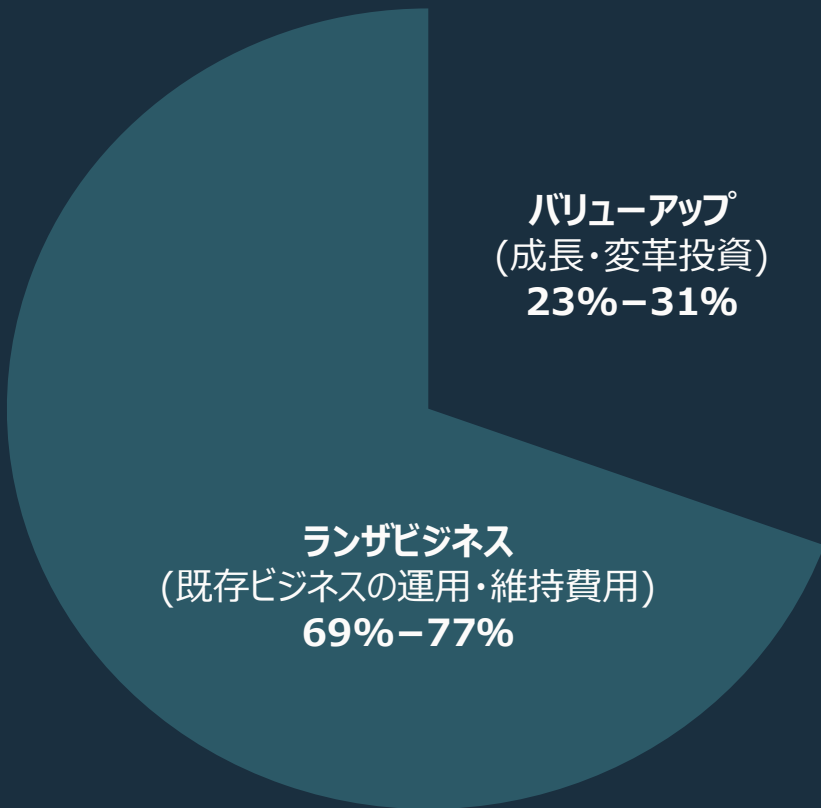
私の経験を元にAI/ML技術の企業実用化によるDX推進に関して**個人的な見解**を共有させていただきます。



AIによるDX推進

ITコスト構造改革の重要性

日本企業のテクノロジー予算分布



0 企業DX戦略

業務変革 社会課題解決 デジタルビジネス創出

1 DXを支えるデジタル技術

AI IoT クラウド

2 DXを支える資産

デジタル人材 データ

3 ランザビジネスの圧縮化

内製化
 FinOps
 クラウドリフト
 マルチクラウド

4 バリューアップ戦略投資

AI
 AIアプリケーション 機械学習
 データ
 プラットフォーム

生成AIの進化背景 — 規模 (スケール) の進化

数十億パラメーター*を超える大規模言語モデル
数兆を超えるデータ要素からの学習
対話型のインターフェイス

* 人工ニューロンとニューラルネットワークから構成されるモデルの係数

生成AIの仕組み

“Hello, how are”



大規模言語モデル



“you”

“Hello, how are you?”

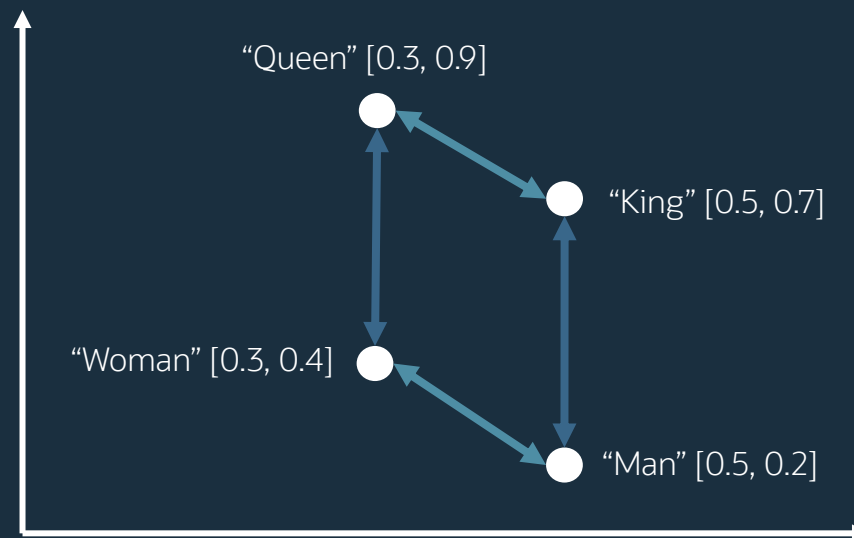
ベクトル埋め込み - 特徴の数値化

“Queen” = [0.3, 0.9]

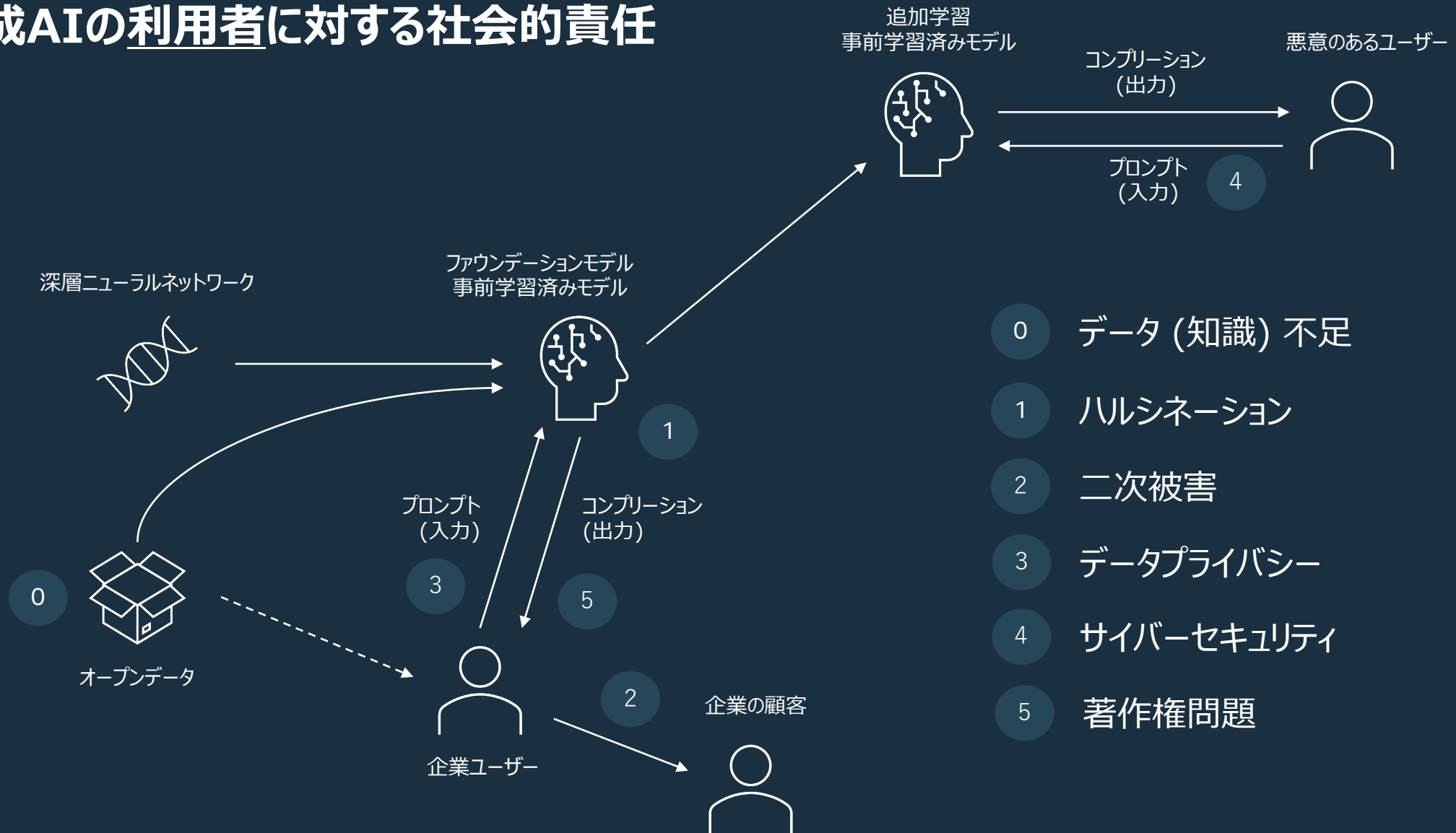
“King” = [0.5, 0.7]

“Woman” = [0.3, 0.4]

“Man” = [0.5, 0.2]

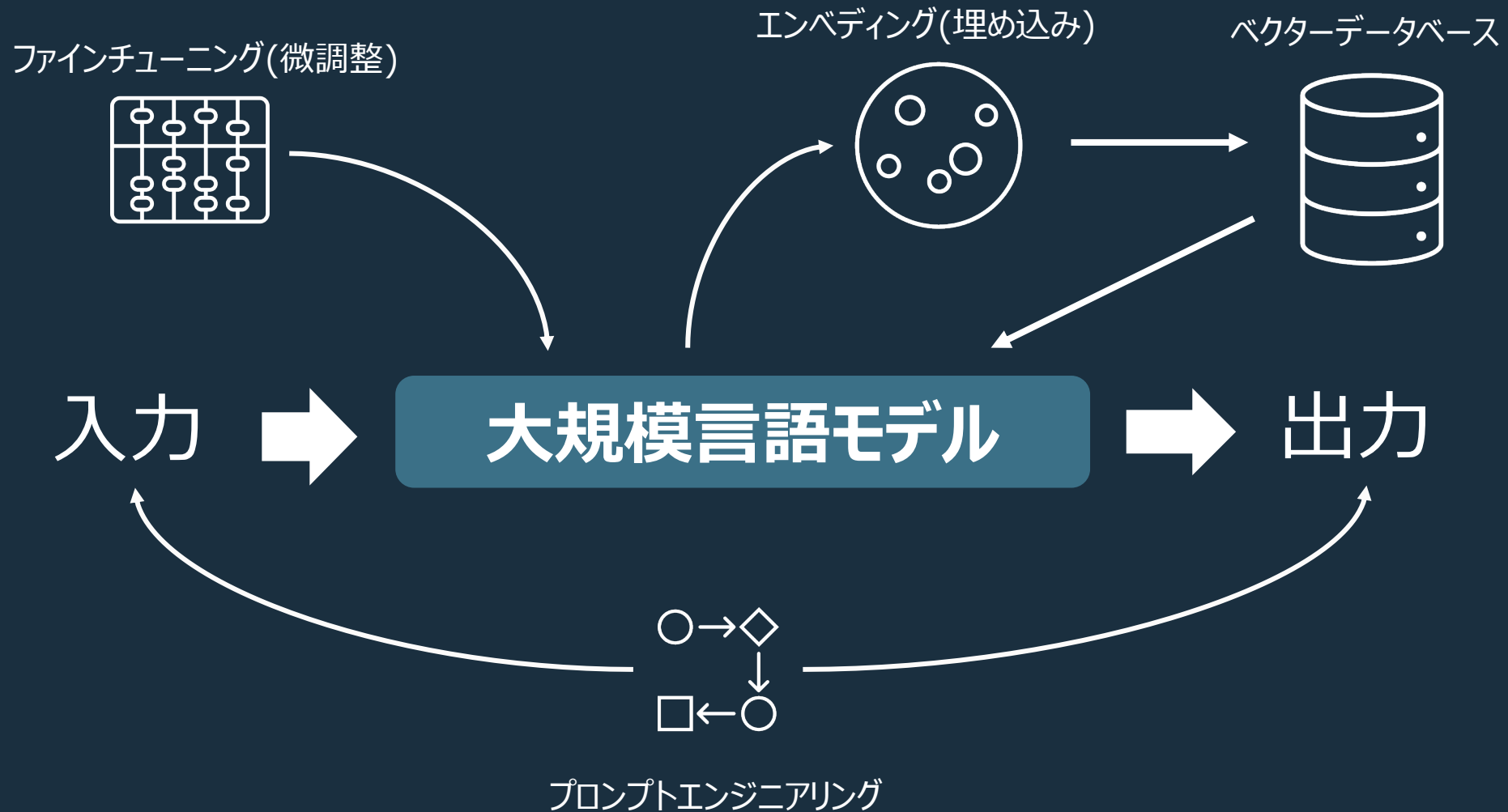


生成AIの利用者に対する社会的責任

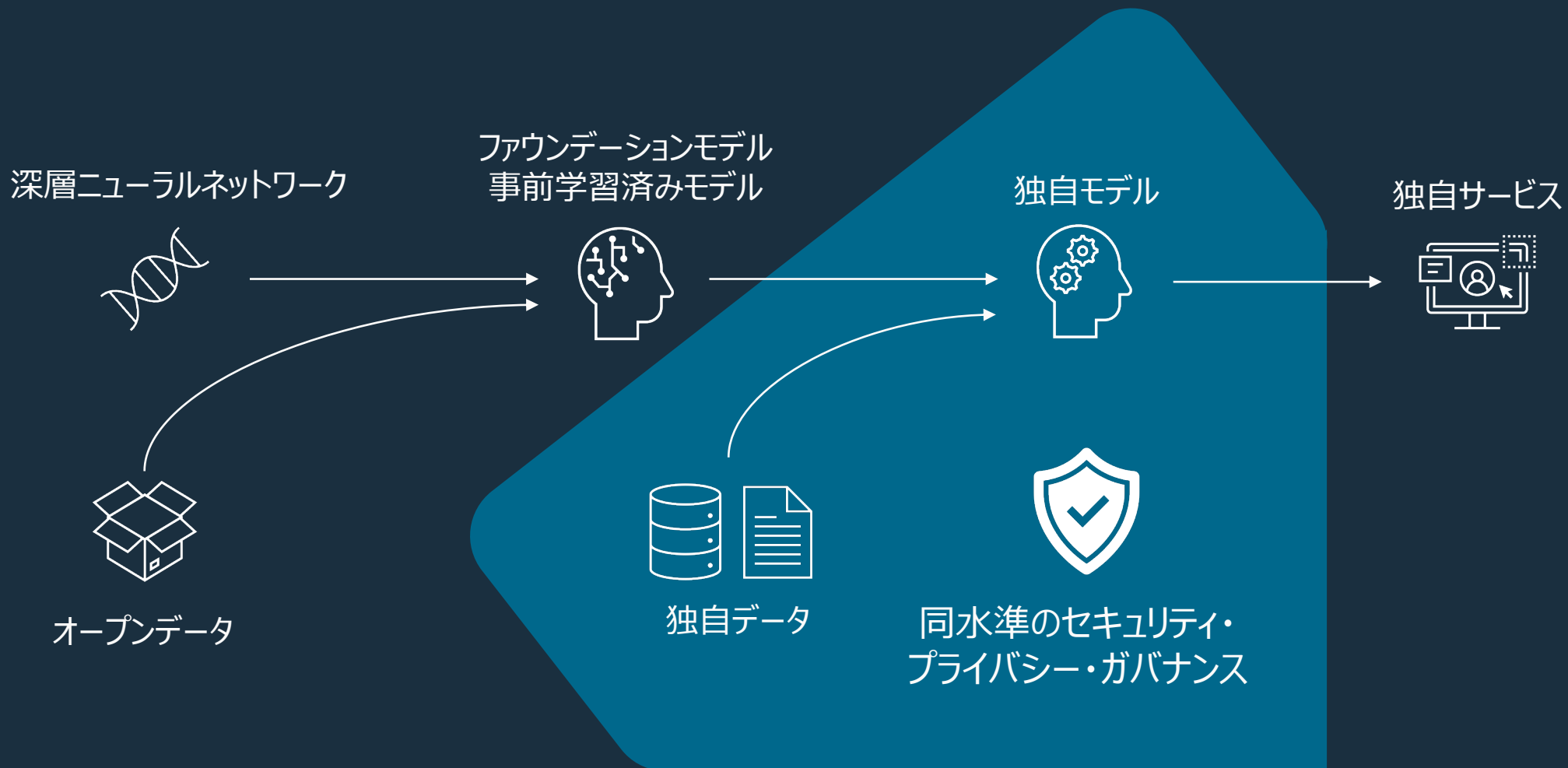


言語モデルに対するアプローチ

手法例

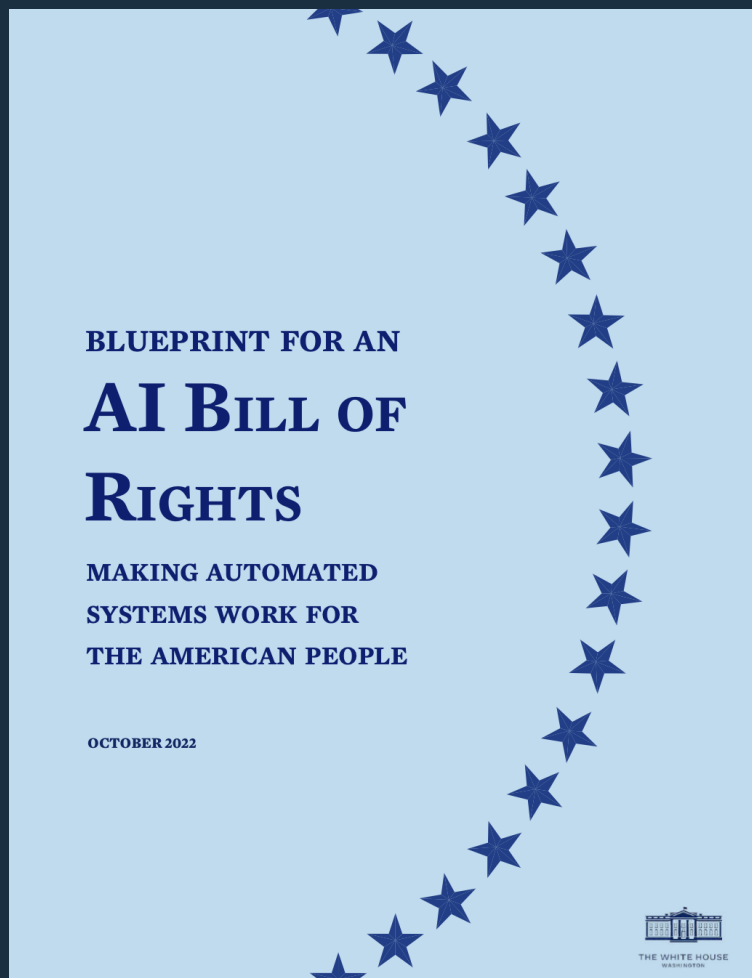


独自モデルに求められるセキュリティとガバナンス



責任あるAIに対するガイドライン例

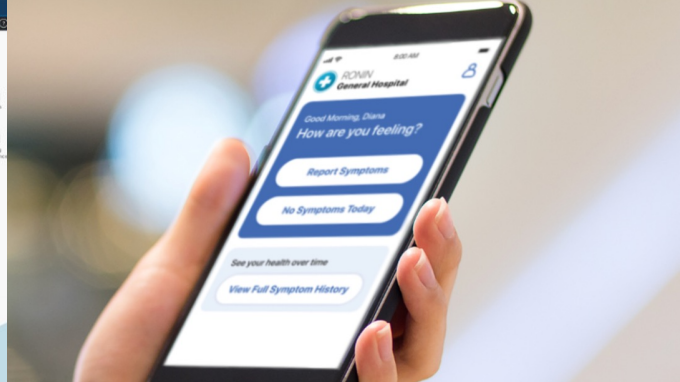
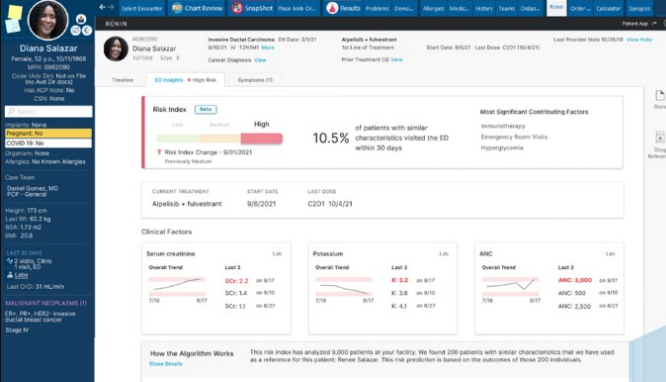
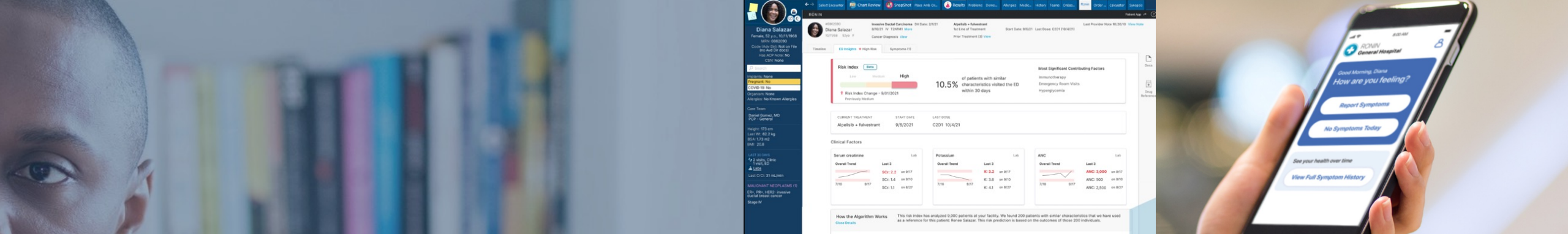
Blueprint for an AI Bill of Rights



AI権利章典（人権保障規定）

- Safe and effective systems (安全で効果的なシステム)
You should be protected from unsafe or ineffective systems.
- Algorithmic discrimination protections (アルゴリズムによる差別からの保護)
You should not face discrimination by algorithms and systems should be used and designed in an equitable way
- Data privacy (データのプライバシー)
You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.
- Notice and explanation (通知と説明)
You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you.
- Human alternatives, considerations, and fallback (人間の代替案、考慮事項、予備)
You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter





RONIN

Analyze and Deliver Patient Data

Take Action on Patient Data

EOM

Capture Patient Data

Electronic Health Record

Care Team

Ronin's Solution: AI System for Predicting Acute Care Usage

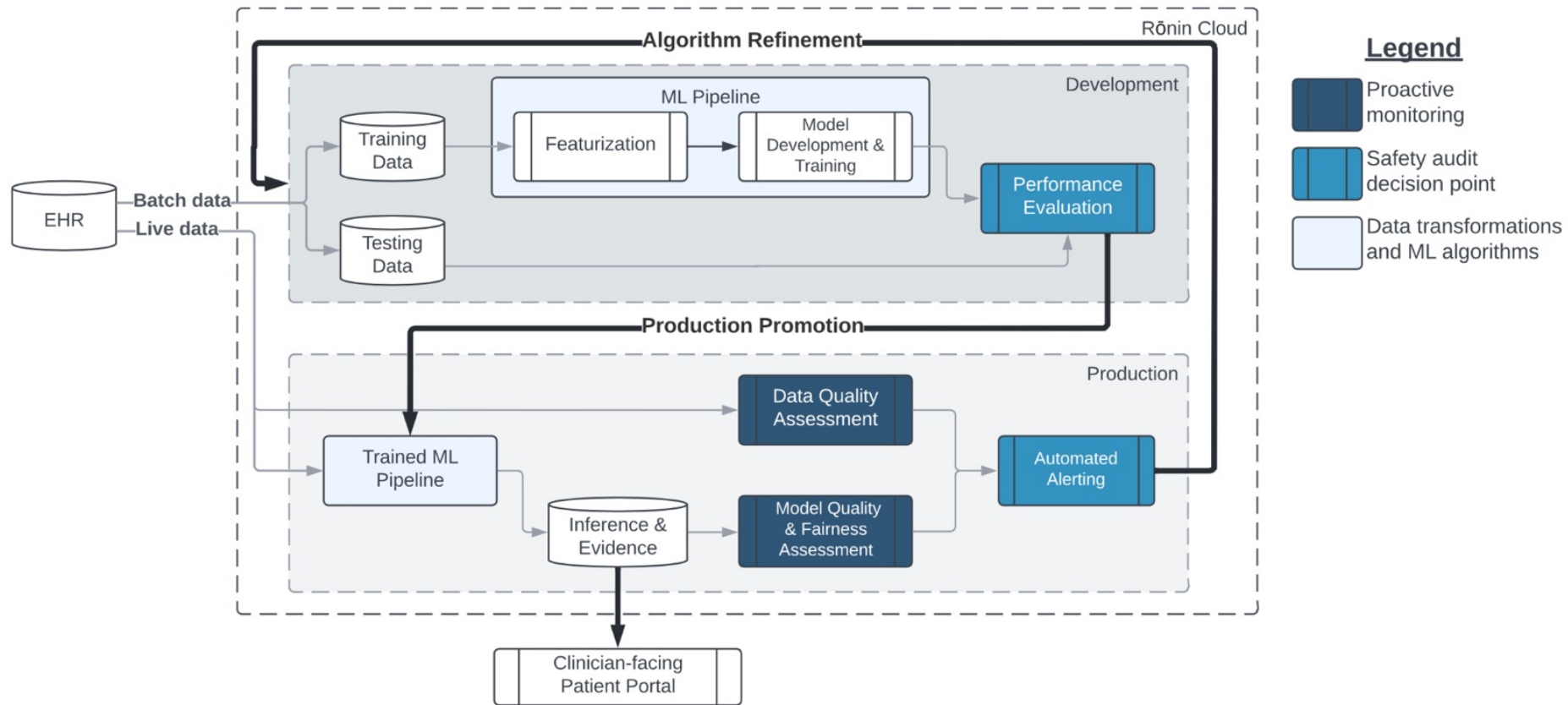
Collaborating with researchers at the top cancer center in the United States, Ronin has developed an advanced AI system that includes a machine learning (ML) model to predict acute care usage in a population of over 28,000 oncology patients. Leveraging 141 routinely-collected data elements from the electronic health record, the model requires no manual data entry or chart abstraction by clinicians. These data elements encompass patient demographics, staging, labs, comorbidities, medications, symptoms, social determinants of health, and prior healthcare utilization including insights from clinical notes extracted using NLP (Natural Language Processing). By analyzing this comprehensive dataset, Ronin's AI system can accurately forecast which patients are most likely to require acute care services in the near future.

<https://www.projectronin.com>



責任あるAI – Project RONIN

<https://www.projectronin.com/blog/revolutionizing-oncology-care-with-artificial-intelligence-predicting-acute-care-usage-in-cancer-patients>



医療業界に於けるAI実用化に向けたガイドライン

<https://blogs.oracle.com/ai-and-datascience/post/improving-cancer-care-ai-bill-of-rights>

- 人工知能による意思決定支援システムの早期臨床評価のための報告ガイドライン
<https://www.bmj.com/content/377/bmj-2022-070904>
- 医療意思決定支援ソフトウェア 産業界およびFDAスタッフのためのガイダンス
<https://www.fda.gov/media/109618/download>
- 医療機器開発のための優れた機械学習の実践： 指針
<https://www.fda.gov/media/109618/download>
- 英国における医療機器の今後の規制に関する協議
<https://www.gov.uk/government/consultations/consultation-on-the-future-regulation-of-medical-devices-in-the-united-kingdom/chapter-10-software-as-a-medical-device>
- 医療画像処理における安全でプライバシーを保護した連携機械学習
<https://www.nature.com/articles/s42256-020-0186-1>
- AI RISK MANAGEMENT FRAMEWORK
<https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development>
- 医療向けの信頼できる AI 導入ガイダンスと保証の青写真
https://www.coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf
- 医療機器としてのAI/MLベースのソフトウェアに対する医薬品の規制枠組みの提案
<https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>

透明性を持った説明

<https://blogs.oracle.com/ai-and-datascience/post/improving-cancer-care-ai-bill-of-rights>

How the Algorithm Works

[Close Details](#)

This model has analyzed 9,000 patients at your facility. We found 200 patients with similar characteristics that we have used as a reference for this patient: Renee Salazar. This prediction is based on the outcomes of those 200 individuals.

Largest Impact on Risk

[102] attributes were evaluated in this risk index. The 10 attributes with the largest negative impact on this patient's risk are shown in this chart.

	Decrease	Risk	Increase
Immunosuppressants			●
Number of previous ED visits			●
Population Transportation Access			●
Recent Symptom (Nausea/Vomiti...			●
Diabetes (Uncomplicated)			●
Hypertension			●
T Stage (Pathologic)			●
Days since last ED visit			●
Race			●
Recent Symptom (Fever)			●

Patient Comparison Breakdown

How this patient compares to 200 patients with similar characteristics.

- Overall: 82% Match
- Age: 95% Match
- Cancer Type: 80% Match
- Medication Treatment Cycle: 80% Match





データセット

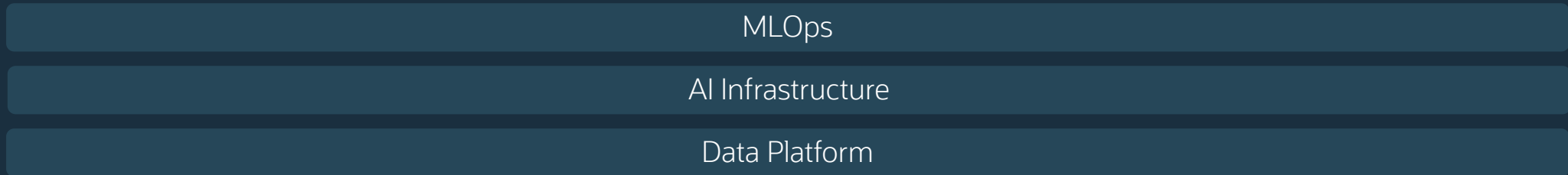
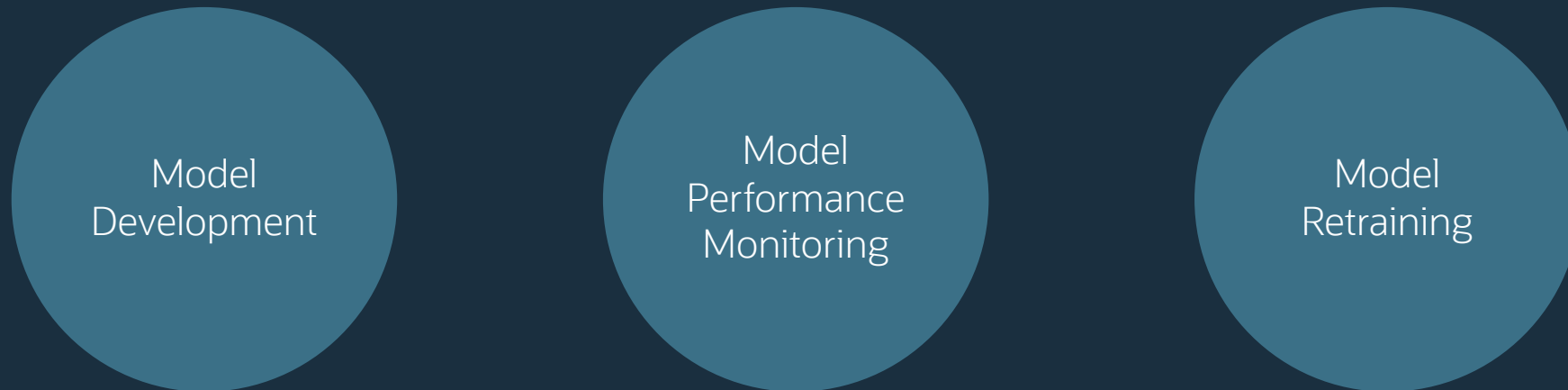
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10287624/>

- テキサス大学MDアンダーソンがんセンターの協力のもと、28,369名・184,138件の電子カルテデータを受領
- 学習データと評価データを4:1の比率で分類。一個人のデータが学習データと評価データに混在しない様に配慮
- 米国保健福祉省 (HHS) の「Healthy People 2020」で概説されている定義を使用し、郵便番号と国勢調査のデータから社会的決定要因 (SDOH) を算出
- 99種のデータ要素から222の特徴を抽出
- 実データをモデル評価のシミュレーションデータとして活用し、現実世界に於けるモデル動作として検証
- リアルタイムデータを継続的に監視する事により動的な医療環境に於けるモデルパフォーマンスの維持を実現

安全、公平、効果的な AI モデルの構築

<https://blogs.oracle.com/ai-and-datascience/post/improving-cancer-care-ai-bill-of-rights>

Ronin は、安全、公平、効果的な機械学習システムの開発と提供に取り組んでいます。これを達成するために、彼らは 3 つの側面からのアプローチに従っています。まず、モデル開発中に、高いパフォーマンスを確保し、敏感な人口統計上のサブグループ間の偏りを防ぐために、広範な検証が実施されます。次に、モデルのパフォーマンスを継続的に監視して、パフォーマンスの低下やドリフトを迅速に特定します。最後に、問題が発生した場合、Ronin は根本原因分析、更新されたデータ、モデルの再トレーニングを通じて問題を迅速に修正します。モデル開発、モニタリング、修正のこの入念なアプローチにより、AI モデルが正確で偏りのない状態を維持するシステムが構築され、臨床医とテクノロジーの間の信頼が促進されます。



社会的な課題を解決する為のデータ特徴抽出

<https://www.healthcareitnews.com/news/how-ai-and-nlp-can-boost-care-remote-patient-monitoring>

Q. Where does natural language processing fit into the whole process?

A. Health data is extremely complex. It's structured in some sense, but very poorly structured in another sense, and it's evolved over time.

One thing we still find is that, no matter how much effort we put into standardizing the structure of the data, is that EHR vendors still have differences in how the data is shared. Also, the percentage of data in the EHR that is actually plain text is astonishing, and we have seen publications that put that number above 80%.

Using natural language processing to be able to extract features out of that data, even structure that data so that it can be used in machine learning models, is critical to obtaining information from that data. Natural language processing is really a necessary step in order to bring structure and meaning to the data that we have in the medical record.

Q. 自然言語処理はプロセス全体のどこに当てはまりますか？

A. 健康データは非常に複雑です。ある意味では構造化されていますが、別の意味では構造化が非常に不十分であり、時間の経過とともに進化しています。

私たちが依然として感じていることの 1 つは、データ構造の標準化にどれだけの努力を払ったとしても、EHR ベンダーによってデータの共有方法に依然として違いがあるということです。また、EHR 内の実際にプレーン テキストであるデータの割合は驚くべきものであり、その数値が 80% を超えているという出版物も目にしました。

自然言語処理を使用してデータから特徴を抽出し、機械学習モデルで使用できるようにデータを構造化できることは、データから情報を取得するために重要です。自然言語処理は、医療記録にあるデータに構造と意味を与えるために本当に必要なステップです。



要素技術のご紹介 | エンベディング (埋め込み)

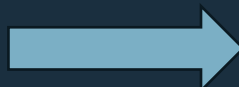
入力

1	朝起きると急な頭痛を感じる事がある
2	大体夕方になると頭痛がする
3	立ちくらみを頻繁に起こす
4	職務中に腰が常に痛い
5	ランニング中に目眩を感じた
6	仕事中に常に目の疲れを頻繁に感じる
7	なかなか寝付けない



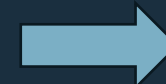
大規模言語モデル
(Cohere Multilingual)

ベクトル
埋め込み



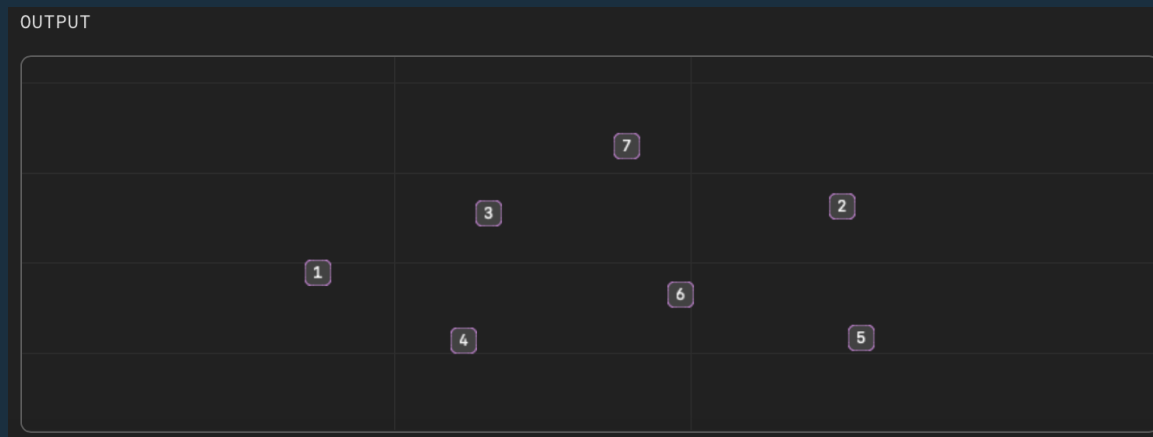
二次元類似性表現

	v_1	v_2	v_3
1	0.23	0.88	0.23
2	0.6	0.3	0.6
3	0.1	0.88	0.88
4	0.12	0.3	0.88
5	0.88	0.88	0.3
6	0.3	0.3	0.1
7	0.4	0.4	0.12



...

ベクターデータベース



ベクトル活用例 | 分類

cohere playground

DASHBOARD DOCUMENTATION PLAYGROUND COMMUNITY

Generate Classify Embed Summarize BETA

View code Save Share

Examples FINTECH EMAIL CLASSIFIER INSURANCE FAQ CLASSIFICATION ARTICLE TITLE CLASSIFIER See more

TRAINING EXAMPLES or Upload a file

Example	健康上の症状	精神上的症状
× 定期的な頭痛がする	<input checked="" type="checkbox"/>	<input type="checkbox"/>
× ぐるぐる目が回るようなめまいを感じる	<input checked="" type="checkbox"/>	<input type="checkbox"/>
× 運動中に急に腰が痛くなった	<input checked="" type="checkbox"/>	<input type="checkbox"/>
× 体がフワフワするようになる目眩を感じる	<input type="checkbox"/>	<input checked="" type="checkbox"/>
× ちょっとした腰の痛みがずっと続く	<input type="checkbox"/>	<input checked="" type="checkbox"/>
× 業務中に強い頭痛が急に現れる	<input type="checkbox"/>	<input checked="" type="checkbox"/>
× 育児中に首から肩にかけての緊張を感じる	<input type="checkbox"/>	<input checked="" type="checkbox"/>
× 最近、感情のコントロールができない	<input type="checkbox"/>	<input checked="" type="checkbox"/>
× 職場で体調が悪くなる事が多い	<input type="checkbox"/>	<input checked="" type="checkbox"/>

MODEL embed-multilingual-v2.0

PARAMETERS

TEXT

- × 1. 朝起きると急な頭痛を感じる事がある
- × 2. 大体夕方になると頭痛がする
- × 3. 立ちくらみを頻繁に起こす
- × 4. 職務中に腰が常に痛い
- × 5. ランニング中に目眩を感じた

OUTPUT

Value	Confidence Level
1. 朝起きると急な頭痛を感じる事がある	<input checked="" type="checkbox"/> 精神上的の症状 61%
2. 大体夕方になると頭痛がする	<input checked="" type="checkbox"/> 精神上的の症状 62%
3. 立ちくらみを頻繁に起こす	<input checked="" type="checkbox"/> 健康上の症状 52%
4. 職務中に腰が常に痛い	<input checked="" type="checkbox"/> 精神上的の症状 77%

CONTROL PANEL

RUN

CLEAR

ラベル付例

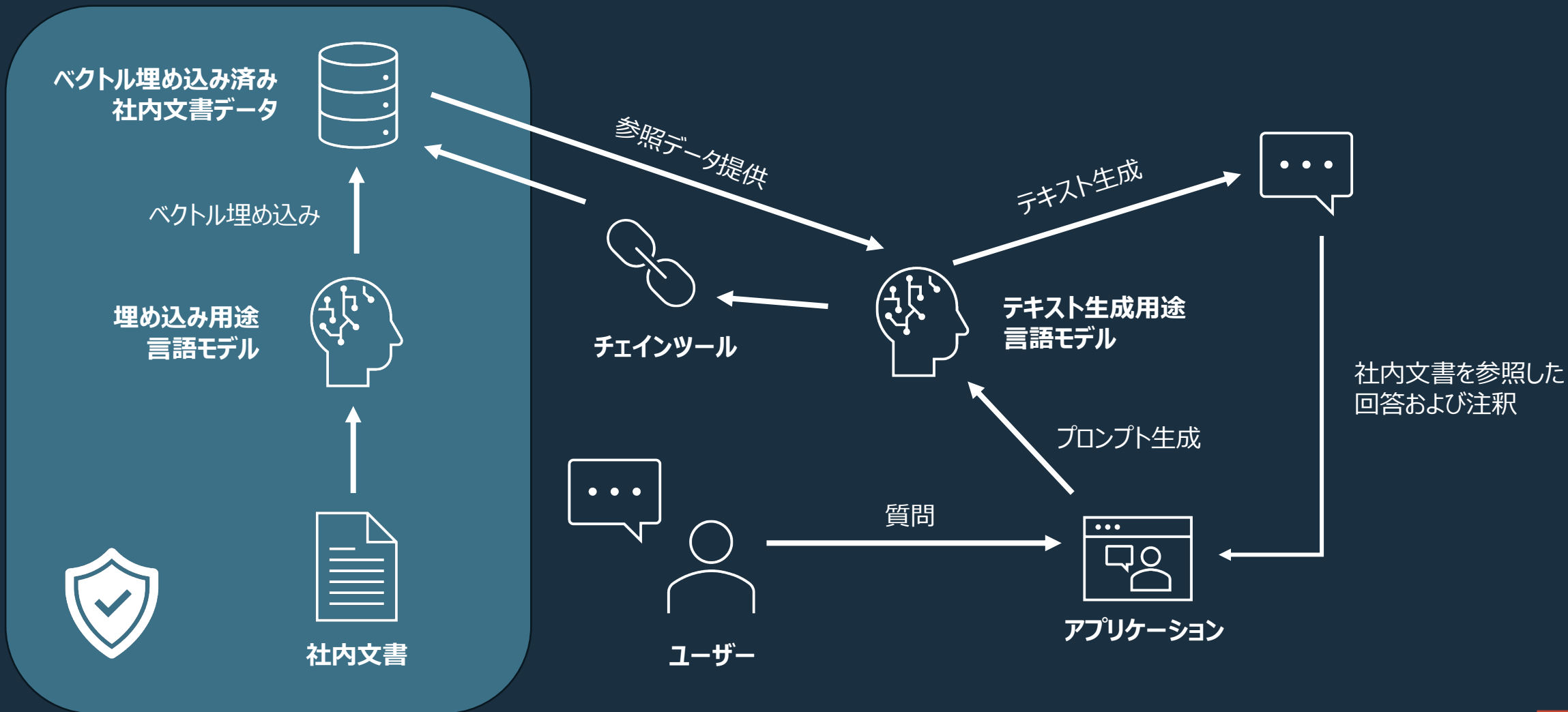
マルチ言語対応
エンベッドモデル

分類するテキスト

分類結果



応用シナリオのご紹介 | RAGを活用した社内FAQサービス



応用シナリオのご紹介 | Citation (引用) の重要性

質問

日本で一番高い山は？

回答

日本で一番高い山は、国の中央部に位置する富士山です。富士山は富士五湖地域に属する火山であり、人気の観光地^{1,2}です。日本人にとっても聖地であり、巡礼者もよく登る山³です。

この山は東京から西に約 100 キロ¹ (60 マイル) のところにあり、晴れた日には市内からも見ることができます。ハイキングや登山の目的地としても人気³です。

富士山の頂上まではいくつかの異なるルートがあり、ルートと個人の体カレベルに応じて、登山には数時間から数日かかる²場合があります。最も人気のあるルートは富士宮口登山道で、富士宮五合目からスタートし、登山に約 2 時間かかる全長 5.5 km (3.4 マイル)^{1,2,3} のトレイルです。

注釈

1. 日本の地形

www.yourprivatesite.com/mountain/info.html

日本の山地・山脈。山地はいくつかの山が集まったもの。山脈はいくつかの山々が長く連なり、帯状に伸びる山地。本州の中央部は「日本の屋根」とよばれる、標高 ...

2. 日本の主な山岳標高

www.yourprivatesite.com/mountain/detail/experience.html

「日本の山岳標高一覧 -1003山-」は、2万5千分の1地形図などの基本測量の成果を基礎として、日本の主な山1003山の山名、最高地点の位置・標高など ...

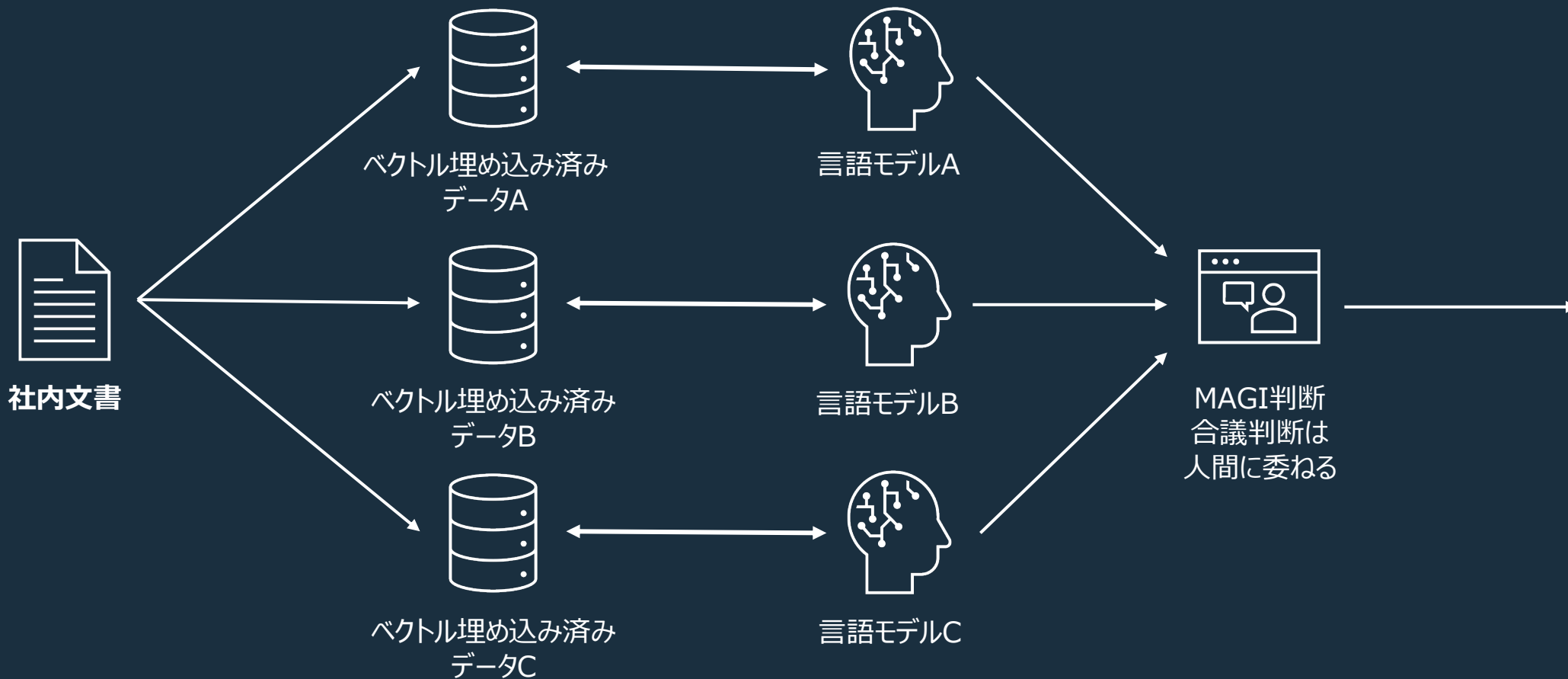
3. 山が多く森林にめぐまれた国土

www.yourprivatesite.com/mountain/detail/feedback.html

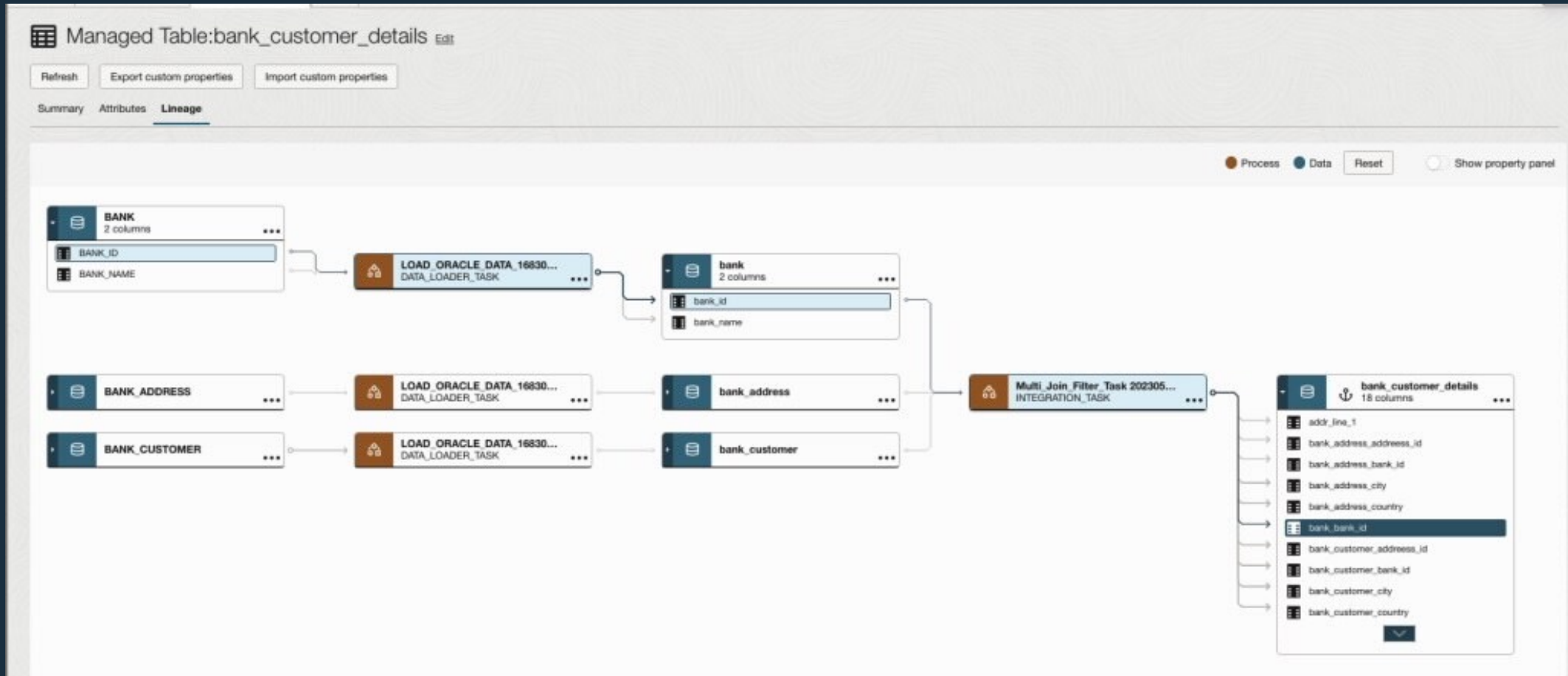
日本の国土はどんな地形をしているでしょうか。山地が多く、大きな川が流れる平野、山に囲まれた盆地（ぼんち）、森林がゆたかな高原。日本の山や森林について調べ ...

個人的な期待値 | MAGIシステム

<https://note.com/fladdict/n/n106b9ce8f7d4>

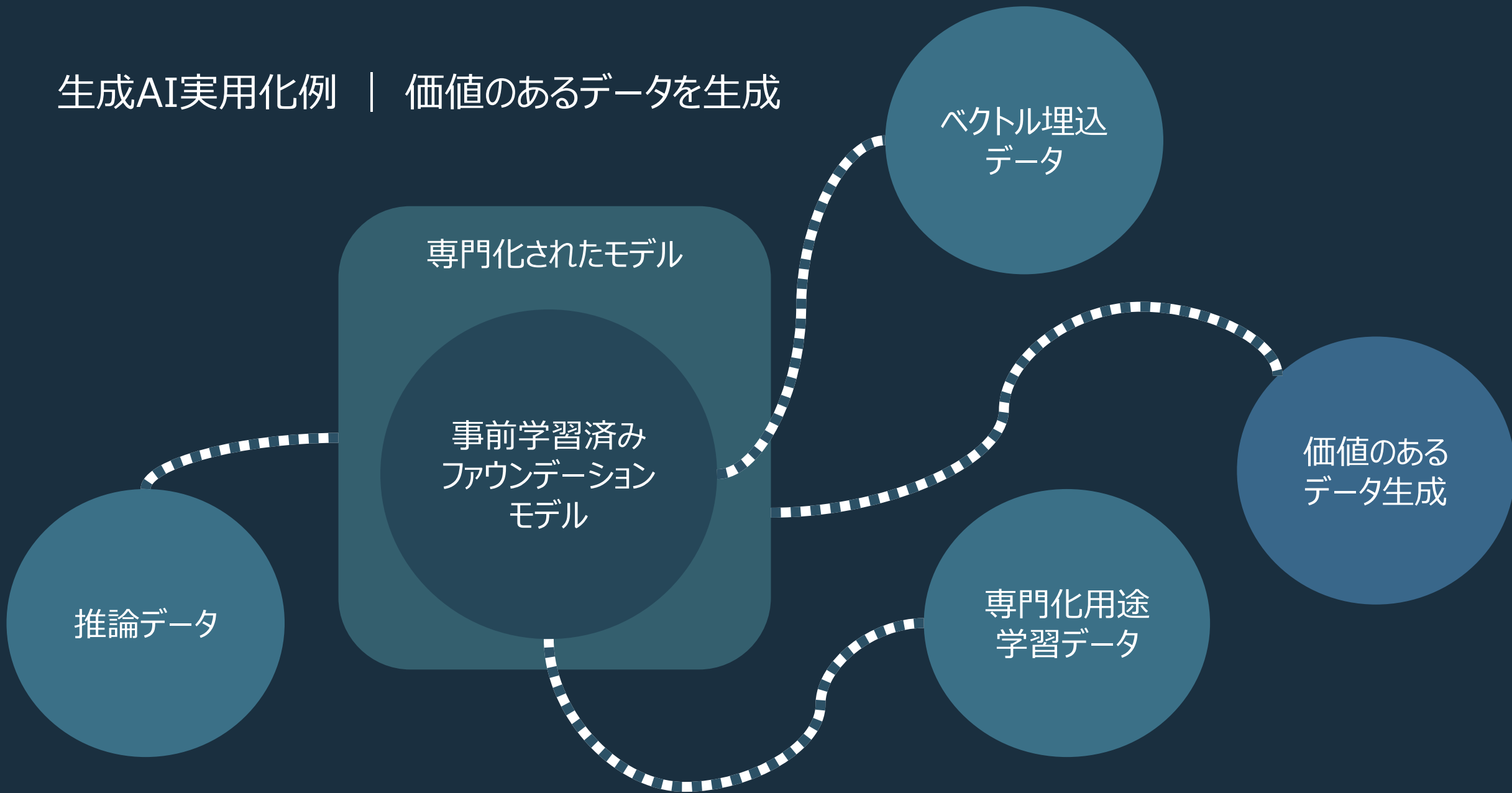


著作権課題 | Data Lineage (データ系列)



<https://blogs.oracle.com/cloud-infrastructure/post/where-did-the-data-come-from-data-lineage>

生成AI実用化例 | 価値のあるデータを生成



AIの要素技術例

Natural Language Processing

- Text Classification
- Sentiment Analysis
- Key Phrase Extraction
- Translation
- Language Detection
- Named Entity Recognition
- Semantic Parsing
- Personal Indefinable Information
- Named Entity Recognition
- Question Answering
- Casual Language Modeling
- Key Value Extraction
- Embedding

Computer Vision

- Image Classification
- Document Classification
- Object Detection
- Optical Character Recognition
- Depth Estimation
- Image Segmentation

Tabular

- Anomaly Detection
- Scoring
- Forecasting

Multimodal

- Speech to Text
- Image to Text

AI Services

API経由で利用、アプリケーションにすぐ組み込み可能



Document Understanding
テキストおよび表の抽出



Language
センチメント分析やキー・フレーズ抽出など、高度なテキスト分析を実行



Hello...

Digital Assistant
チャットボットなど、対話を通じた問い合わせ応答や、タスクの実行



Speech
音声の転送、字幕の作成、メタデータの生成



異常検出
ビジネス・データの望ましくないイベントまたは観測をリアルタイムで識別



Vision
イメージ内のオブジェクトの検出および分類



Forecasting
今日のトレンドから明日何が起こるか予測



私たちのミッションは、人々が
新たな方法でデータを理解し、
本質を見極め、無限の可能性を
解き放てるよう支援していくこ
とです。



顧客企業とのDX新規事業共創を行う専門チーム ソーシャル・デザイン推進本部による無償支援

以下の開発初期フェーズのプロセスを無償支援します



アイデアの創出

ワークショップや
ディスカッションを通して、
社会課題解決につながる
既存事業の拡張を模索



サービスの開発

オラクル社内エンジニアが、
外部のアカデミアや
パートナーの先進技術も
活用してMVPを開発



事業化の推進

同様の課題解決を図る
他企業連携や既存にて
施策展開されるフィールドを
活用して社会実装を推進

ORACLE JAPAN

FY24

**FOCUS
MEASURES**

Cloud for Japan



日本のためのクラウドを提供

AI & ML for Customers



お客様のためのAIを推進